Steffens, P. R. (1998). "Applying Diffusion Models with Regional Heterogeneity," *Marketing Letters*, 9(4): 361-369

Title: Applying Diffusion Models with Regional Heterogeneity

Author: Dr Paul R Steffens

Lecturer

Graduate School of Management

The University of Queensland

Brisbane QLD 4072

AUSTRALIA

Accepted for publication in *Marketing Letters.*

Key Words: Diffusion Models; Sales Forecasting; Sales Models.

Word Count: 3799 (Main Body)

Author's Biographical Note

Dr Paul Steffens is a lecturer in marketing at the Graduate School of Management, University of Queensland, Australia. His research interests are sales modeling and new product development.

Applying Diffusion Models with Regional Heterogeneity

Abstract

Recent studies of innovation diffusion have investigated cross-country heterogeneity, but implicitly assumed within-country homogeneity. As such, these studies potentially overlook within-country variations in diffusion patterns, which may be even more important to marketing managers and researchers alike. The current paper is concerned with such intra-country variations using one of many possible *a priori* segmentation schemes, namely geographic segmentation. It empirically demonstrates that when substantial regional variations in diffusion patterns occur, taking account of these regional differences improves both short- and long-term forecasting under certain conditions. Regional differences in diffusion patterns also provide some important normative implications.

Key Words: Innovation Diffusion Models; Sales Forecasting; Pricing and Advertising Policies.

1. Introduction

Innovation diffusion models provide an important tool for business managers and researchers to study, forecast or manage the sales growth of a new product or service. With few exceptions, published applications of diffusion models are at a national level (Mahajan, Muller and Bass, 1993). A number of recent studies have examined inter-country heterogeneity of diffusion patterns (see review by Parker, 1994). However, the notion of intra-country variations in diffusion, which may be even more important to managers and scholars alike, has been overlooked.

In the current paper, we are concerned with geographic segmentation as one example of an *a priori* segmentation scheme to explore intra-country variations in diffusion. This notion has been addressed in the general diffusion literature (Rogers, 1995) and modeling literature (Mahajan and Peterson, 1979; Gore and Lavaraj, 1987; Weerahandi and Dalal, 1992). The purpose of the current paper is to empirically demonstrate the potential shortcomings of diffusion studies at a national level in terms of their descriptive, forecasting and normative purposes.

The paper presents the findings of an application of a diffusion model for a new agricultural input product. Analysis of four agricultural regions reveals that distinctly different diffusion patterns exist. We investigate the implications for forecasting national sales generated by analyzing the variations in the regional diffusion patterns. Finally, we present some preliminary normative implications of the regional diffusion analysis through an illustration of optimal pricing and advertising strategy.

2. Empirical Study

We present the results of an empirical study involving the diffusion of a new agricultural input chemical. This product represents a significant innovation in farm practice for a particular crop. We define the model used in the study and present the estimation and forecasting results.

2.1. The Diffusion Model

The model is based on the Bass (1969) model

$$\frac{dN(t)}{dt} = \left(m - N(t)\right)\left(p + \frac{q}{m} N(t)\right) \tag{1}$$

where $N(t)$ is the cumulative number of adopters at time t, m is the upper limit of adopting population, p is the coefficient of external influence, and q is coefficient of internal influence.

To use the Bass model in the context of sales of a non-durable product, the amount of product purchased by each consumer must also be modeled. The simplest and most common assumption is that sales are simply proportional to the number of adopters (Lilien, Rao and Kalish, 1981; Norton and Bass, 1987). For the current application, product sales are truly annual. The product is normally applied to the crop only once a year. To model sales, we assume that the diffusion process (1) occurs in continuous time, and that sales in each year depend on the cumulative number of adopters at that time. A further modification to the model is required since, historically, there is a base level of sales for a similar product used for a different application. While this similar product is not considered relevant to the diffusion process under investigation, the available sales data do not distinguish between the two products. Hence, the model of annual sales is given by

$$S_i = r\,N(i) + S_0 \tag{2}$$

where S_i is sales in year i, r is average annual sales per adopter, and S_0 is the base level of sales attributable to a similar product. If $N(0) = 0$, the closed form solution of Equations (1) and (2) is

$$S_i = K \frac{1 - e^{-(p+q)\,i}}{1 + \frac{p}{q} e^{-(p+q)\,i}} + S_0 \tag{3}$$

where $K = r\, m$.

2.2. Fit to National and Regional Sales Data

The diffusion model is fitted to sales data, which is normalized by dividing by area of crop planted in a given year to correct for the impact of weather. Parameters of Equation (3) are estimated using nonlinear least squares (NLS), as originally advocated for the Bass model by Srinivasan and Mason (1986). The product under investigation is used within four distinct agricultural regions, each accounting for between 20% and 35% of crop production. To study the regional nature of the diffusion, Equation (3) was fitted to the normalized sales data for each region and the national level. Initial inspection of the data revealed that in Region 2, normalized sales are abnormally low in two years that are severely drought affected. In these years, normalizing sales by area planted does not fully correct for the extreme impact of weather. Hence, a dummy variable for "severe drought" was introduced for Region 2 and the national level for these 2 years.

The parameter estimates (with approximate error estimates in parentheses) and fit statistics, R^2, mean squared error (MSE) and mean absolute error (MAE), are reported in Table 1, and the fit is plotted in Figure 1. Some estimation difficulties warrant comment. Figure 1 shows clearly that diffusion in Regions 3 and 4 is not well advanced. The difficulty of using early sales data (prior to the inflection point) for estimation of the ultimate sales potential, K, for diffusion models is widely recognized (Mahajan, Muller, and Bass, 1993). Preliminary analysis revealed that for Regions 3 and 4, unconstrained estimation resulted in non-sensible parameter estimates for K. Fortunately, the more advanced regions provide ideal analogies to guide judgmental parameter estimation. Furthermore, management was comfortable estimating the maximum "usage rate" for the product. Based on their judgment, K is constrained to 21.0 for

Regions 3 and 4. These constraints had a very small impact on the resultant model fit for Regions 3 and 4, confirming that *K* cannot be reasonably estimated from the data alone.

The most important finding of the regional analysis is the widely varying patterns of diffusion in the four regions. Diffusion in Regions 1 and 2 is substantially more advanced than in Regions 3 and 4. Diffusion in Region 3 is occurring rapidly, but was slow to start (low external influence, high internal influence), whereas Region 4 exhibits an almost opposite pattern, with a slow, but steady, diffusion (high external influence, low internal influence).

2.3. Forecasting

We investigate whether the above identification of the regional variations is of benefit when forecasting national sales. Alternatively, forecasting may be impaired, since slower diffusing regions (Regions 3 and 4 for this application) require a longer period of data before reasonable parameter estimates can be made.

Preliminary analysis identified that there is very little difference in short-term forecasting via the regional and national data for most years. However, two conditions are identified when regional analyses are of substantial benefit for forecasting. First, long-term forecasting is enhanced when the fastest diffusing region has passed its point of inflection but the national diffusion pattern has not. Here, the upper limit for this fast diffusing region provides an ideal analogy to estimate upper limit for other areas. Second, short-term forecasting is enhanced when diffusion "takes off" in one region. In this instance, a regional analysis can forecast the rapid short-term sales growth reasonably well. However, the impact of this regional sales growth gets masked in the aggregate data, as these regional sales are still small relative to other regions.

The benefit of the regional analysis for long-term forecasting is best elucidated by considering the forecasting of the upper limit of sales, *K*, in years 11 or 12. Examination of

Figure 1 reveals that diffusion is substantially more advanced in Region 1 than at the national level. Table 2 compares both the forecasts of the upper limit of sales, *K*, and its asymptotic standard error, for Region 1, Region 2, and the national level in years 11 and 12. Comparing Region 1 with the national level, it is clear that the estimates for *K* are fairly close. However, the standard errors are substantially smaller for Region 1. Similarly, the standard error is smaller in Year 12 for Region 2. As a consequence, considerably more confidence can be placed on the estimates for Region 1, and to a lesser extent on those for Region 2. Conversely, however, estimates of *K* for the slower diffusing Regions 3 and 4 are considerably worse if based on the data alone. Fortunately, the early diffusing region, Region 1, provides the ideal analogy to estimate the upper limit, *K*, for Regions 3 and 4. Hence, overall for Years 11 and 12, long-term forecasting is improved by a sensible analysis of the regional diffusion patterns.

The benefits of the regional analysis for short-term forecasting are best illustrated when forecasting the last year of sales in year 11. In this year, sales in Region 3 have recently "taken off," but this effect gets masked in the national data, because Region 3's sales are still a small fraction of overall sales. Table 3 presents a comparison of this short-term forecast for the aggregated regional analysis and the national analysis. Two forecasting situations are considered. The first is an unconstrained fit based entirely on the data. The second case assumes the upper limit of sales is estimated externally. For this case, the value of *K* for each forecast was constrained to the value shown in Table 1. In both forecasting cases, the aggregated regional analysis provided a substantially better forecast than the national analysis. In fact, the error is approximately halved. The difference between the forecasts can be attributed to the recent and rapid growth in Region 3. Figure 1 reveals that while this impact is evident in the regional data, it gets masked in the national data. Hence, the national level forecast substantially underestimates the one step ahead forecast in this situation. We conclude that, under conditions such as these, the analysis of regional diffusion improves short-term forecasting.

3. Normative Illustration

A normative illustration is presented to demonstrate some implications of regional heterogeneity in diffusion. No general normative results are possible since an infinite combination of regional diffusion patterns is possible. Therefore, the illustration presented is based on the empirical illustration. Where possible, parameter values are taken from the above results. Other nominal parameter values are based on previous empirical studies.

We base the illustration on the models first introduced by Robinson and Lakhani (1975) for pricing, and by Horsky and Simon (1983) for advertising. These are modified for our discrete time application for a non-durable product. We consider the monopolistic optimal pricing strategy, P_i, and advertising strategy, A_i, over a finite planning horizon. Specifically, we consider the maximization of the profit function

$$\pi = \sum_{i=1}^{N} \left[\left(P_i - vc \right) \cdot S_i - A_i - fc \right] \cdot \left(1 - d \right)^{-i} \tag{4}$$

where vc is the variable cost of producing the product, fc is the fixed costs of production, and d is the discount rate (above inflation).

We include both the influence of price on the adoption rate and the volume of product purchased per adopter. Two different functional forms for the influence of advertising on sales are illustrated. Model 1 follows the assumption originally made by Horsky and Simon (1983), that advertising influences only the coefficient of external influence, p. Alternatively, Model 2 reflects the recent findings of Bass, Krishnan and Jain (1994), that advertising does not have a differential effect on p and q. It is assumed that advertising has no impact on the volume of product purchased per adopter. The model of sales is:

$$S_i = \left(r_0 \, e^{-\gamma P_i}\right) N_i \tag{5}$$

$$N_i = m \, \frac{1 - p_i \left(1 - N_{i-1} / m\right) / \left(p_i + q_i N_{i-1} / m\right) e^{-(p_i + q_i)}}{1 + q_i \left(1 - N_{i-1} / m\right) / \left(p_i + q_i N_{i-1} / m\right) e^{-(p_i + q_i)}} \tag{6}$$

where N_i is the number of adopters in year i, and p_i and q_i are now functions of price and advertising. r_0 and γ are parameters of the model. As before, we let $K_0 = r_0 \, m$. The specific functional forms for p_i and q_i are:

Model 1: $$p_i = \left(\alpha + \beta \ln \overline{A_i}\right) e^{-\delta \overline{P_i}} \qquad q_i = \theta \; e^{-\delta \overline{P_i}} \tag{7}$$

Model 2: $$p_i = \left(\alpha + \beta \ln \overline{A_i}\right) e^{-\delta \overline{P_i}} \qquad q_i = \left(\theta + \beta \ln \overline{A_i}\right) e^{-\delta \overline{P_i}} \tag{8}$$

where α, β, θ and δ are parameters, $\overline{P_i}$ is the normalized price ($P_i - P_{base}$) and $\overline{A_i}$ is the normalized advertising ($(A_i / A_{base} / K_0) + 1.0$).

A simulation based on the empirical example was conducted to determine the optimal pricing and advertising strategy for both models. Specifically, Equation (4) is maximized using equations (5), (6), and either (7) or (8) to determine sales. It is assumed that advertising levels can vary between regions, but to comply with regulatory requirements, price must be identical in each region. Several parameters are set based on the empirical fitting: namely $K_0 = K$, $\alpha = p$, $\theta = q$. Hence, if all $P_i = P_{base}$ and $A_i = 0$, then the diffusion patterns are identical to the empirical ones. We set $P_{base} = 1.0$, $vc = 0.67$, $fc = 300$, and $A_{base} = 10$. We impose the constraint that $P_i \geq vc$, implying that the product cannot be sold below the marginal cost of production. The parameters relating to the quantitative influence are set to nominal values that are consistent with past empirical studies (Kamakura and Balasubramanian, 1988; Bass, Krishnan and Jain, 1994;

Kalish, 1983). Specifically, $\gamma = 0.5$, $\delta = 0.5$ and $\beta = 0.7$.

The optimal strategy is determined for the regional analyses and compared to that for the national level. The national level pricing and advertising strategy is then applied to the regional analyses (i.e. the annual pricing and advertising levels are used in Equations (4), (5), (6), and either (7) or (8) with the parameters from the regional analyses). As such, this determines the impact on profit of treating the national diffusion as homogeneous, rather than regional, when developing pricing and advertising strategies.

Figure 2 presents the results of the normative illustration for Model 2. It compares the optimal pricing and advertising strategies derived for the regional analyses with those for the national analysis. Table 4 lists the impact on profit for each case. It is clear that, for all cases, the optimal pricing policy is monotonically increasing, while optimal advertising is an initial burst type strategy. While not plotted, the results for Model 1 are very similar to Model 2, except that the initial burst of advertising decays more rapidly. In fact, advertising is almost zero after year 1. This is expected, since advertising influences only the innovation coefficient in Model 1.

The more important comparison, of course, is between the national and regional analyses. Both analyses result in similar optimal pricing strategies, though the price increases more rapidly for the regional analysis than the national analysis. However, the primary difference between the two analyses is the level of advertising expenditure across the regions. In particular, the regional analysis suggests higher levels of advertising than the national analysis for Regions 3 and 4, but lower for Regions 1 and 2. Also, the initial burst of advertising for the regions decays more slowly than for the national level in all regions except Region 4. Ultimately, this results in a profit improvement of 12.1% for Model 1 and 4.8% for Model 2. We conclude that the regional analysis substantially improves normative pricing and advertising guidelines for this illustrative case.

4. Concluding Remarks

The paper empirically demonstrates that the accepted practice of applying diffusion models at a national level may result in some important regional differences being overlooked. An understanding of regional variations in diffusion has an immediate descriptive benefit. The study also reveals that sensible analysis of the regional variations in diffusion aids both short- and long-term prediction of national sales under certain conditions. A simulation study demonstrates some normative implications of regional heterogeneity. Finally, regional segmentation is just one example of *a priori* segmentation. Heterogeneous diffusion patterns across other segments of the population warrants further investigation.

5. References

Bass, F.M. (1969). "A New Product Growth Model for Consumer Durables." *Management Science*, 15(5) (January), 215-227.

Bass, F.M., T.V. Krishnan and D.C. Jain. (1994). "Why the Bass Model Fits Without Decision Variables." *Marketing Science*, 13(3) (Summer), 203-223.

Gore, A. P. and U. A. Lavaraj. (1987). "Innovation Diffusion in a Heterogeneous Population." *Technological Forecasting and Social Change*, 32, 163-168.

Horsky, D. and L.S. Simon. (1983). "Advertising and the Diffusion of New Products." *Marketing Science*, 2(1) (Winter), 1-17.

Jain, D.C. and R.C. Rao, (1990). "Effect of Price on the Demand for Durables: Modeling, Estimation and Findings." *Journal of Business & Economic Statistics*, 8(2) (April), 163-170.

Kalish, S. (1983). "Monopolist Pricing with Dynamic Demand and Production Cost." *Marketing Science*, 2(2) (Spring), 135-159.

Kamakura, W.A. and S.K. Balasubramanian. (1988). "Long-Term View of the Diffusion of Durables: A Study of the Role of Price and Adoption Influence Processes Via Tests of Nested Models." *International Journal of Research in Marketing*, 5, 1-13.

Lilien, G.L., A.G. Rao, and S. Kalish. (1981). "Bayesian Estimation and Control of Detailing Effort in a Repeat Purchase Diffusion Environment." *Management Science*, 27(5) (May), 493-506.

Mahajan, V., E. Muller and F.M. Bass. (1993). "New Product Diffusion Models." In J. Eliashberg and G.L. Lilien (eds.), *Handbooks in Operations Research & Management Science, Vol. 5: Marketing*, (349-408). Amsterdam: North-Holland.

Mahajan, V. and R.A. Peterson. (1979). "Integrating Time and Space in Technological Substitution Models." *Technological Forecasting and Social Change*, 14, 231-241.

Norton, J.A. and F.M. Bass. (1987). "A Diffusion Theory Model of Adoption and Substitution for Successive Generations of High-Technology Products." *Management Science*, 33(9) (September), 1069-1086.

Parker, P.M. (1994). "Aggregate Diffusion Forecasting Models in Marketing: A Critical Review." *International Journal of Forecasting*, 10, 353-380.

Robinson, B. and C. Lakhani. (1975). "Dynamic Price Models for New Product Planning." *Management Science*, 10 (June), 1113-1122.

Rogers, E.M. (1995). *Diffusion of Innovations* (4th ed.). New York: The Free Press.

Srinivasan, V., and C.H. Mason. (1986). "Nonlinear Least Squares Estimation of New Product Diffusion Models." *Marketing Science*, 5, 169-178.

Weerahandi, S. and S.R. Dalal. (1992). "A Choice-Based Approach to the Diffusion of a Service: Forecasting Fax Penetration by Market Segments." *Marketing Science*, 11(1) (Winter), 39-53.

	Parameter Estimates				Fit Statistics		
	p	***q***	***K***	S_0	R^2	**MSE**	**MAE**
Region 1	6.23E-04 (1.46E-03)	0.919 (0.345)	18.89 (1.46)	3.06 (0.87)	98.0%	0.75	0.74
Region 2	3.41E-03 (5.10E-03)	0.799 (0.294)	27.18 (2.95)	1.23 (1.52)	94.4%	5.19	1.79
Region 3	5.23E-05 (1.01E-04)	0.859 (0.194)	21.00 (fixed)	1.54 (0.49)	93.9%	1.00	0.79
Region 4	4.33E-03 (3.99E-03)	0.393 (0.108)	21.00 (fixed)	3.51 (0.66)	95.9%	0.47	0.54
National	1.20E-02 (6.68E-03)	(0.389 (0.161)	22.40 (5.88)	1.32 (0.81)	95.9%	1.08	0.88

Table 1: Parameter Estimates and Fit Statistics for Model Fitted to Normalized Sales

	Upper Limit of Sales *K*		Asymptotic Standard Error	
	Year 11	**Year 12**	**Year 11**	**Year 12**
Region 1	17.7	18.9	1.3	1.5
Region 2	26.7	27.2	2.3	3.0
National Analysis	17.7	18.3	2.2	5.9

Table 2: Comparison of Long-Term Forecasting for Normalized Sales

	Unconstrained Fits			Constrained Fits		
	Forecast Error	**Squared Error**	**Relative Error**	**Forecast Error**	**Squared Error**	**Relative Error**
Region 1	-1.87	3.49	-9.6%	-0.04	0.00	-0.2%
Region 2	-3.40	11.58	-28.9%	-3.54	12.55	-30.1%
Region 3	2.53	6.41	18.1%	1.23	1.51	8.8%
Region 4	-1.97	3.87	-13.9%	-2.31	5.33	-16.3%
Aggregated Regions	-1.21	1.47	-7.8%	-0.91	0.83	-5.9%
National Analysis	-2.89	8.33	-18.6%	-1.66	2.76	-10.7%

Table 3: Comparison of Short-Term Forecasting (Year 12) for Normalized Sales

	Model 1	**Model 2**
National Analysis	$247	$312
Regional Analysis	$277	$327
Improvement	12.1%	4.8%

Table 4: Comparison of Profits

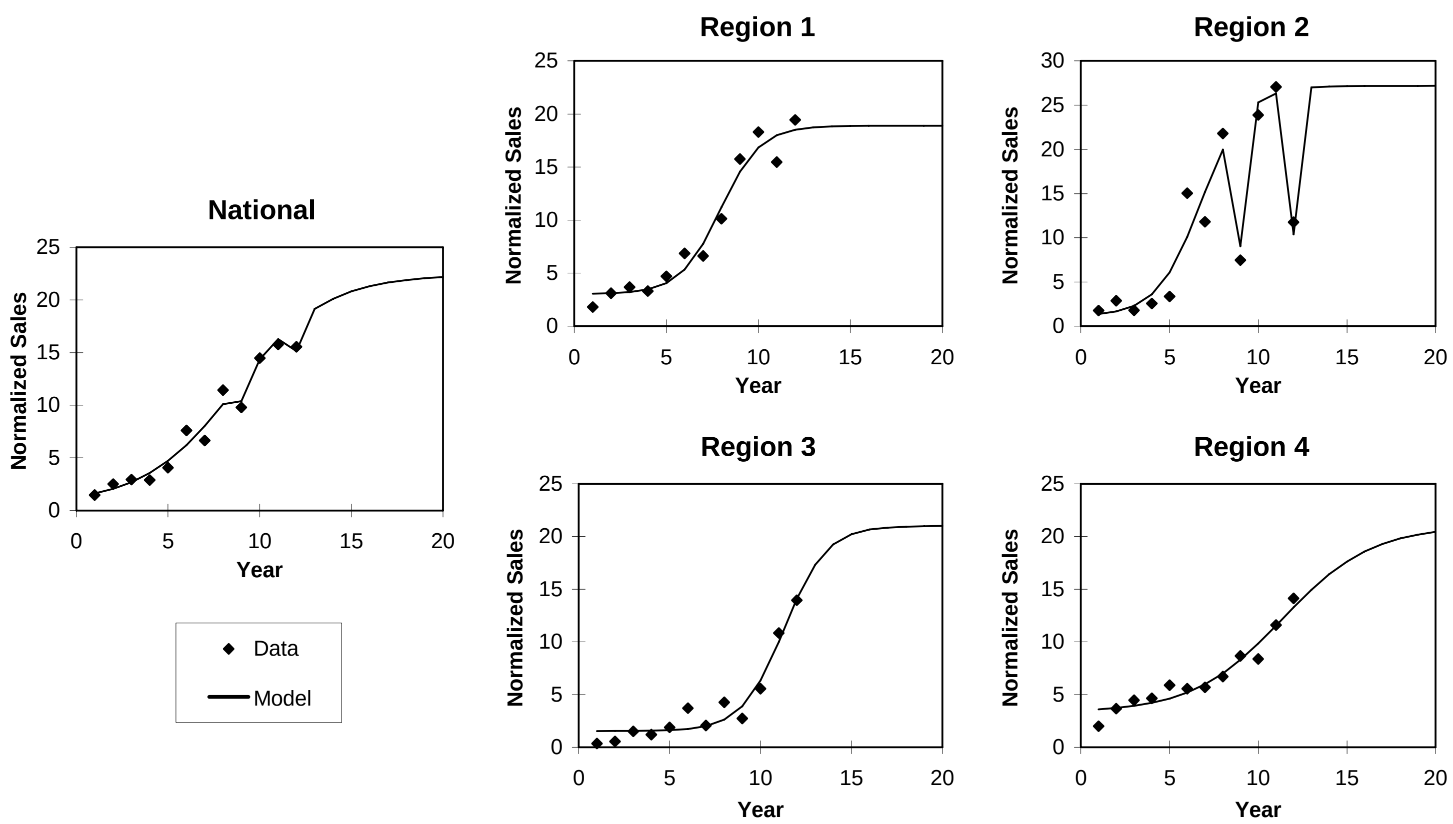

Figure 1: Fits and Forecasts of Regional Normalized Sales

ɪ

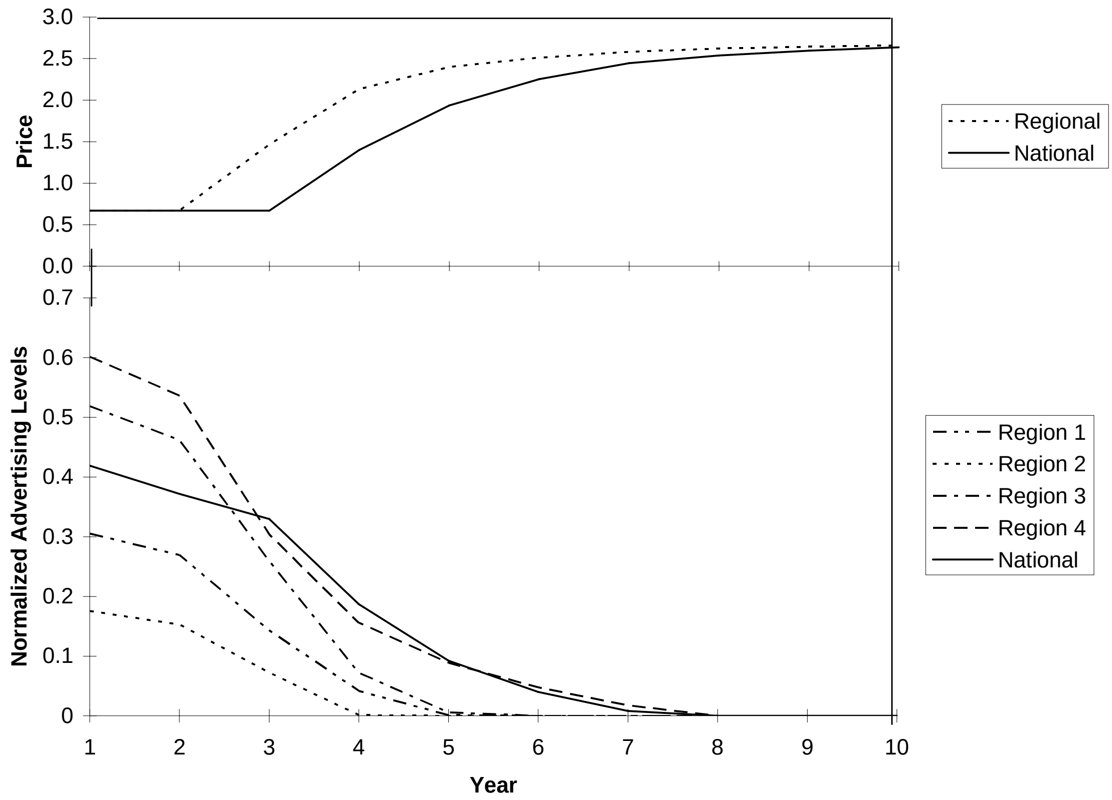

Figure 2: Comparisom of Optimal Pricing and Advertising Policies